INTER-ORGANIZATIONAL DATA INTEGRATION PROBLEMS

Frank G. Goethals, Monique Snoeck, Wilfried Lemahieu, Jacques Vandenbulcke,
Department of Applied Economics - K.U.Leuven
Naamsestraat 69, B-3000 Leuven, Belgium
{Frank.Goethals, Monique.Snoeck, Wilfried.Lemahieu,
Jacques.Vandenbulcke}@econ.kuleuven.be

ABSTRACT

In this paper, we discuss eight high-level problems that show up when partnering companies decide to share information. It is important in practice to be aware of the existence of these problems and to deal with each of these problems. The problems concern the fact that it is difficult to identify what information flows to automate, that companies have a different viewpoint upon boundary objects, that they have to agree on a data format with an appropriate functionality, that investments have to be distributed among the parties, that service levels should be agreed upon, that partners should preserve the value of the data sharing, that a data owner is needed at an inter-organizational level and that partners may change over time.

Keyword: Business-to-Business integration, integration problems,

INTRODUCTION

For many years companies have been optimizing their internal functioning. With the advent of the Internet, however, information sharing among companies has become much more feasible so that the optimization effort is nowadays mostly no longer confined to that of individual enterprises: optimization can happen at the level of Extended Enterprises. Clearly, Extended Enterprises that are successful in their Extended Enterprise wide optimization effort will undercut other value chains that fail to do so and only optimize the individual components of the value chain rather than the value chain in its entirety.

The optimization of the functioning of Extended Enterprises relies heavily on information sharing because of two main reasons: (1) as information is being shared data inconsistencies across enterprises can be eliminated so that all companies possess an accurate picture of reality to act upon, and (2) as new information is being shared, new business practices become possible. As we know from the process-paradigm, companies can execute a task if they are in some state. Nowadays ICT makes it possible to transmit information on a multitude of states in real-time. For example, in the past only two states were recognized in the ordering process: ‘order placed’ and ‘delivery received’. Nowadays one is informed that an order was received alright, has been accepted, has been planned, that the goods have been picked, that the goods have been loaded in the truck, that the truck has left, etc.

Let us illustrate the two main reasons for information sharing with two examples.

- One important problem that is discussed extensively in Supply Chain Management literature is that of the Bullwhip-effect, the fact that variability in demand is magnified as one moves up the supply chain (i.e., from reseller, over seller to manufacturer). Case studies have proven that through VMI (Vendor Managed Inventory) the Bullwhip-effect can be strongly reduced (see e.g. [1]). Doing VMI requires an intensive sharing of stock and sales data

among different companies in the supply chain. The information sharing is in this case so intensive that it is unrealistic to assume that the same amount of information could be shared and processed manually, without directly connecting the computer systems of the different companies.

In discussions on VMI it is often stated that the product data of the seller and the reseller needs to be synchronized before the VMI effort can be started, and that the product data should be kept synchronized during the VMI effort [2]. Indeed, typically both the seller and the reseller will have stored product data internally before getting into VMI. In [3: p11] it is stated that '30% of items in retail catalogues have data errors' and that 'for new products it can take up to four weeks for complete and accurate item data to reach the retailer for entry into their procurement systems.' Clearly, the product data that is stored in the systems of the seller and the reseller may be out of synch. If data on the wrong products is available, or if wrong data is available on the right products VMI cannot work.

Furthermore, the seller that is involved in a VMI effort may start storing information about the reseller's stock. The seller will receive information on the reseller's sales and knows how many units he has delivered to the reseller. From this, he can calculate the stock that is supposed to be at the reseller's site. Still, in practice that calculated number often seems to differ from the actual stock at the reseller's site. The reseller then has to transmit stock level data to get valid data at the seller's site.

- As another example, companies can pursue Product Lifecycle Management (PLM). Through PLM companies can get better products faster to the market, can provide better support to the customers, etc. [4]. PLM requires the collaborative creation, management, dissemination, and use of product definition information across the Extended Enterprise from concept to end of life of the product [5]. Product-related data is traditionally created and stored in many applications. Customer requirements can be defined in a CRM system, an 'engineering BOM' of the product can be designed in a CAD system (Computer Aided Design), and a 'manufacturing BOM' is used in an ERP solution to manage stocks of products and semifinished products, etc. Of course, the engineering BOM may differ from the manufacturing BOM as the latter is adapted to the way the product is to be manufactured [6]. The manufacturing BOM may still be different from the way a specific product is manufactured in practice. Some materials may for example be replaced by other materials. Data on the product that is actually produced then also should be stored so as to enable a decent after-sales service. Although all this data is stored in different systems (which are in an Extended Enterprise typically located in different companies), people that use one system may need data that is created in another system. For example, if the manufacturing company is not to produce products that have been replaced or modified in the engineering company, information on such changes should be communicated. Before starting a collaboration, the product definition information is typically scattered over the different partners' existing systems. To be able to develop, produce, and maintain a qualitative product, partners need to share the information they own.

  In the PLM context we see that the introduction of additional information flows enables new practices. While it is technically feasible to design

semifinished products in isolation, purely on the basis of the specification of a desired ‘interface’, it is more valuable to design semifinished products in consultation. Many different versions of requirements and designs may have to be sent back and forth. Also, CAD/CAM systems of different designers of different parts of an assembly can communicate so as to detect whether the change in the dimension of the design of one part results in the violation of spatial constraints (i.e., bumps with the current design of another part) [7]. The idea to regularly transmit information between different designers is clearly valuable.

Partners have or want to have knowledge about the same object, but initially data about that object are disseminated or even duplicated across different isolated systems so that necessary data may not be accessible and that inconsistencies can arise. Moreover, partners may not be aware of the fact that data about this object are stored elsewhere too.

The **challenge** for companies within an Extended Enterprise is then the following:

> Companies should (only) have access to valid information on which they would agree they need to have access to.

In what follows we investigate a number of lower level ‘information sharing problems’ that are likely to show up when setting up an infrastructure for data sharing. When we introduce the problems we will illustrate them with the VMI and the PLM practices mentioned above. Furthermore, sometimes we will illustrate some problem with a case study we have done. One case study that will be mentioned repeatedly is the Tradcom-case. Tradcom [http://www.tradcom.com] is a company that offers a marketplace for companies in the BeNeLux (Belgium, the Netherlands, and Luxembourg) to trade indirect goods and services. The suppliers have a long term relationship with Tradcom (actually, the suppliers commonly own Tradcom), and the integration between suppliers and Tradcom is quite tense. At the customer side the coupling is less stringent. Because it concerns indirect goods, not making up the core business of the customers, customers do not want to invest too much money in linking their systems directly to the platform.

We note that companies in an Extended Enterprise may not only want to share ‘explicit knowledge’ but also ‘tacit knowledge’ (as defined in [8]). When we were investigating existing B2B standards during our research we found that industry standards so far have focussed on the exchange of explicit, structured and semi-structured data (i.e., “records” [9]). How to share tacit knowledge is typically not the topic of such standards. This perception is in line with what Nonaka [10] found for the internal working of isolated companies. Nonaka states that it is often overlooked in Western companies that creating knowledge depends on [p21] “tapping the tacit and often highly subjective insights, intuitions, and ideals of employees” rather than on simply mechanistically processing objective information. The means for making use of such knowledge, and for articulating the tacit knowledge, often take the form of slogans, metaphors, and symbols. Computer systems cannot deal with metaphors and the like; only people can. Therefore, when it comes to tacit knowledge computer systems get a supporting role. In [11] and [12] it is for example reported that online chat or instant messaging systems successfully support tacit knowledge sharing. The software proposed in an experiment in [13] was used to locate experts and to communicate across the Internet. In that case,

the chat dialogues were saved in a database as records of tacit knowledge sharing. This is one way to start making the tacit knowledge explicit[1].
Although companies in an Extended Enterprise may want to share tacit knowledge, we only investigate the exchange of explicit knowledge in this paper. Tacit knowledge is (at least till some moment in time) not stored in computer systems and has as such a number of distinct characteristics which make it inappropriate to treat explicit and tacit knowledge as one and the same.

PROBLEMS THAT SHOW UP WHEN TRYING TO SHARE DATA

A number of important problems show up when companies want to share information. In what follows eight problems (derived from [16, 17, 18, etc.]) are presented. Each time we

- define the problem,
- show how it is related to the challenge defined above,
- discuss the relevance of the problem, especially in the context of the Extended Enterprise, and
- illustrate the problem in the context of the practices presented above: VMI and PLM.

The problems we discuss are the following.

- First and foremost, companies have to define what information flows are valuable from a business point of view: what data does a company want to use, when should it get that data for the data to be useful, etc. Identifying what information flows to automate is not an easy task.
- When one wants to identify information flows, one will stumble across another problem: if companies want to share data about an object, they have to acknowledge that they may use the object in different tasks and that they, therefore, may have another viewpoint on the object. They need thus to map their viewpoints before they can actually go about sharing data.
- Only once companies have dealt with the two previous problems they know what data they want to share. The next step is then to identify how the data should look like: partners have to specify in what form they will share information.
- To realize the information sharing, investments will need to be made. Partners have to agree on who will bear costs for installing, maintaining and upgrading the systems.
- Once the investments are made, business continuity can only be ensured if the data is provided by the systems as needed. The systems should thus offer appropriate service levels in terms of availability, response time, etc..
- Furthermore, if data is being shared with a partner, this gives the different partners new sources of power. The data receiver may forward the data to a third party or may inadequately secure his systems, the data provider may not

[1] Kim et al. [14] would use the term ‘implicit knowledge’ in this case, rather than ‘tacit knowledge’. This is their reflection on the difference between (1) the Polanyi classification [15] of knowledge into tacit knowledge and explicit knowledge, where tacit knowledge *cannot* be expressed externally; and (2) Nonaka’s vision [8] where tacit knowledge is knowledge that is *currently not expressed* externally. Kim et al. follow Polanyi, and refer to knowledge that *can* be expressed externally but currently exists internally with the term ‘implicit knowledge’.

pay enough attention to data quality, etc. Partners must preserve the function of the data.
- A party should be designated that can decide what can or has to happen with some data, and what cannot: the data owner. Often it is not clear who is the designated data owner.
- Finally, it has to be recognized that all of the above problems have to be dealt with in the frame of changing relationships. New partners may be added, and former partners may be removed from the network.

In this paper we do not investigate how companies can deal with each of these problems. Rather, we show what those problems are.

Valuable information sharing practices have to be identified

First and foremost, partners have to identify what information sharing practices are valuable. An information sharing practice's value depends upon the data that gets to some party at some moment in time.

PROBLEM DESCRIPTION

Identifying what data should be transmitted electronically when to who is difficult. This is more so if the information flows were not yet existent at organizational level before. Creativity and coordination may be needed to redesign existing information flows and to find new valuable information flows (and ways to realize them). Some party may be required to capture new data internally (e.g., more fine-grained data), and new data content might be identified that only exists at the level of the Extended Enterprise, which needs to be captured somehow too. It has to be identified when the interested party needs the data as this puts restrictions on when he should receive the data *at the latest*. Also, given the type of the data and its usage, the data will have an update frequency. This influences the preference of companies to request for data if they need it or to subscribe for data transmissions and thus to leave the initiative to transmit to another party.

When starting a B2Bi effort it is very likely that the existing information flows between the partners have not been 'architected', let alone that they would have been made explicit. This turns many existing information flows invisible, complicating their digitalization.

RELATIONSHIP TO THE CHALLENGE

Who should be expected to have access to *what* data *when?*

RELEVANCE

In an Extended Enterprise there may be a desire to set up a number of non-standard data exchanges, possibly of data that was previously considered to be highly confidential. If valuable data flows cannot be identified and realized, but your competitors *can* realize and harness them, you drop behind. You fail to identify and to seize the opportunities that are embedded in the good relationship between the companies.

ILLUSTRATIONS

The introduction of EDI enabled companies to transmit existing documents such as purchase orders digitally. This way the traditional replenishment process was automated. Later on, VMI could be established as a substitute for the traditional replenishment process just because new information flows (also realized through EDI) were introduced. Sellers only used to receive (and to have knowledge about) the purchase orders of the resellers. With VMI the sellers regularly receive information

about the resellers' sales orders and stock states [21]. Companies that pursue VMI not only have to agree on *what* messages to transmit, but also on when those messages should be transmitted (only after a request, or regularly as a subscription with a publication once a week, every day, several times a day, etc.).
In the PLM domain it is acknowledged that if different companies are interested in the same product data, information transmissions will need to be specified that ensure that the data the companies use are valid. Very often, new information flows have to be established just to make sure that the existing data are and remain valid.
The idea to regularly transmit information between different designers is clearly valuable and requires many decisions to be made. For example, if someone is reworking an artefact that is being used by others in their decision making process, should a new version be transmitted to the others every 15 minutes, every hour, every day or every week?

The partners have a different viewpoint upon objects

In the previous section we said that companies have to determine what data they want to share. Doing this is complicated by the fact that the partners have a different perception of objects on which they want to share data. This is because they use the object differently.

PROBLEM DESCRIPTION
Knowledge is localized and embedded [22]. Integrating the data about an object is problematic as different user groups have a different perception of the object and are thus interested in partly the same and partly different data. After all, they execute different tasks using the object and different business rules may apply to the object. Still, their views are interrelated and the information in the different views should be aligned.
Successful data exchanges between two parties (and thus also between two enterprises) generally fit with theories on boundary objects (or trading zones). Star & Griesemer [23] defined boundary objects as follows: "Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. […] They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation" [23, p393]. Let us consider the concept 'flight' as an example of a boundary object. For a pilot flying an airplane, it does not matter who is sitting where on his airplane, but it is important to know that his flight is going every day at the same moment. He needs to know the weather conditions for the flight, the tail wind, etc. For the stewards who help people check in it is important to know the right people are put on the right flight the right day, and the cooks need to make sure that the right food is available in the right quantity for a specific flight (e.g., pre-ordered vegetarian food). For the cooks the destination of the flight is not important, nor is the identity of the travellers or the pilots. To them, a flight is a number of dishes.
Levina and Vaast [24] identified two basic requirements for an object to be a boundary object. First, the artefact has to acquire a local usefulness. That is, agents in each field must use and make sense of the artefact in the context of each field. Secondly, the artefact needs a common identity. To make this possible, a joint field (which serves to bridge the separate fields) must be established within which agents jointly recognize and value the artefact.

Relationship to the challenge

The data that is *valid* is determined by many parties together (i.e., several parties influence the state of the object); and all partners need data that is valid to them (and some events that happen to the object may not be relevant to them).

We note that validity not necessarily implies consistency. Data that can be considered by the Extended Enterprise to be valid for one company is not necessarily considered valid for another company. For instance, let us consider an engineering company that decides to forward its entire engineering BOM to a number of suppliers, each of which are only interested in a subpart of this BOM. Updates to the BOM are relevant for one party, but not to another party. That is, without the update the data is not consistent, but maybe still valid for some supplier.

Relevance

The vision some have to integrate enterprises for example to the extent that only one electronic version of an order is saved, and that Purchase Order and Sales Order entirely melt together, is threatened by the idea that different companies interpret the concept 'order' differently (i.e., 'order' is a boundary object). Simply stated, different companies may have a number of common *and* a number of private events that can affect the state of the object. When it comes to the common events, it has generally been acknowledged that the lifecycles of the different perceptions of the boundary object should be compatible. For example, both companies should agree that an order first has to be paid, and then has to be delivered; not the other way around [25]. This does not create an additional burden on the wish to store only one copy of the boundary object. However, when it comes to the private events, the boundary object is brought in a different state for one company, but not in another state for the other company. For example, checking the customer history may bring the order in the state 'urgent handling' or the state 'postponed handling' for the supplier, but leaves it in the state 'PO sent' for the customer. Consequently, not all information on the boundary object can usefully be stored in a single shared electronic object.

In an Extended Enterprise companies are willing to share their view on the object and thus to share private data. Partners recognize the viewpoint of the other party and the rules that are embedded in that viewpoint and they may jointly make changes to this view. Partners in a community 'coproduce boundary objects' [26].

Changes that are made to one view of the object but that are not reflected in other views (while they should be) create inconsistencies and thus invalid data. Improvements (and changes in general) to the object are then neglected.

The data that is shared can be coarse-grained or fine-grained. Storing only the coarse-grained data and not the fine-grained data that lies at its basis could be a bad decision as the fine-grained data is lost in such a case. If new data requirements arise (e.g., formulated by a partner) or if one wants to check the origins of the coarse-grained data the original fine-grained data is needed. Still, "What is typically done in designing a database is to use the desired output […] as the design specification and to set up the database to match." Companies then only store the coarse-grained data because this seems to be the data that is actually needed [27].

Parties should only see the information that is relevant to them and should get a view on the information that is adapted to their function. Information overload can deteriorate their performance.

Illustrations

Carlile [22] studied the four primary functions that are dependent upon each other in the

creation and production of a product: (1) sales/marketing, (2) design engineering, (3) manufacturing engineering, and (4) production. He found that the specialization of knowledge complicates working across functional boundaries and accommodating the knowledge created in another practice.
Clearly different people look differently at product data. Designers make a CAD design file and machine operators consider NC (Numerical Control) files. Users of the created part may only be interested in the outside appearance of the part, and whether that part fits in the space that is foreseen, rather than in the vector graphics of the designer. Consumers of a product may want to see how the product appears in different settings (e.g., a refrigerator placed in different kitchens). Furthermore, an engineering Bill of Materials (BOM) may be changed (for example by adding information about machines that will be used during production) to create a manufacturing BOM. Furthermore, designers need to be aware of some rules about the product in order to do their job decently. For example, 'gas A should not pass in the proximity of component Y if its temperature is above 80 degrees'. For users, knowledge is important about rules such as 'the faucet should be kept at a distance of 150 centimetres from electric points'.
Although all data concerns the very same product, parties are interested in information that is to some extent common and to some extent not.
With VMI we see that for the reseller it is (to some extent) irrelevant how the size of a specific order is determined. The seller has to consider issues such as grouping different orders for delivery, rounding off to the pallet, etcetera. For the reseller the order quantity is just a given number of units he expects to receive, for the seller it is a number that should possess several characteristics.

An appropriate data format has to be defined

Once it is known what information the companies want to share, the partners have to look for a way to share the data so that it offers the desired 'functionality'.

PROBLEM DESCRIPTION
For data to be functional it should be understandable. One big step in understanding the data is to know what the data is about; to give some context to the value that is actually transmitted. For example, it may be stated that some value that is transmitted is a supplier number. Although this already drastically reduces the interpretations that can be given to the data one wants to transmit, it may not be sufficient to give a full understanding of the data: companies also need to agree on how some value should be interpreted within some context. For example, should the date 10/05/2006 be understood as being the tenth of May or the fifth of October?
Finding a data format is not easy. The partners may have data in systems they have developed themselves or that were created by different vendors. These systems are likely to support different data formats and their interfaces may not be documented. Also, even though many XML industry standards have been developed for data exchange, many companies have legacy investments in EDI systems what makes the choice for XML messages less evident.
An important decision companies have to make concerns the annotation and structuralization of messages. Different understandable formats can be suggested for transmitting data. Different formats, however, allow a different functionality. If the data is meant to be 'fully functional' for another computer system, using the data will be greatly facilitated if the data is annotated. That can be done by sending the data in an XML format that follows some (standardized) XML schema. To this purpose, the sender can try to push forward the degree of structuralization, so as to transform

unstructured data into semi-structured data. The other way around, if the sender would not like the user to use the information for unknown (or undesired) purposes, he can transmit data that is not fully functional. He can hamper the receiver by not annotating the data, and sending highly unstructured documents. In this case, data that is highly structured at the sender's site (e.g., prices of products in a relational database) may be transmitted in unstructured documents (e.g., highly graphical brochures). An information owner who wants to prevent poaching and the like (see later) may desire the usage of a format that allows 'less functionality' [28, p104], limiting the value of the transmission for the information receiver.

RELATIONSHIP TO THE CHALLENGE

The partners can only get access to the data if the systems can be enabled to talk to each other.

RELEVANCE

If no common format is available no data exchanges are possible. From the Tradcom case we have studied, it is clear that the internal storage format often matters, especially in Market B2Bi. Tradcom allows its customers (with which it does Market B2Bi) to send orders in the format they use internally (e.g., SAP) so as to lower the burden on the customers for doing business with Tradcom. However, Tradcom does not just accept every internally used format: it is capable of receiving messages from counterparties in *a number of* vendor formats (such as SAP), but it is not capable of understanding messages in some proprietary format of the customer. At the supplier-side, where Tradcom forms an Extended Enterprise, Tradcom-specific XML schemas have been developed. All suppliers have to enable their systems to speak that Tradcom-specific language. Their internal data storage format is thus of no direct relevance.

Companies may want to restrict functionality. As an example, a purchaser who is interested to buy products from a supplier may want to look into the supplier's catalogue. The supplier may not want the purchaser to load the entire catalogue into his system because this would make it too easy for the customer to compare prices with competitors. Still, it is valuable for the purchaser that when he selects the products to buy in the catalogue, that the data on these products is made persistent in his system. A solution to this problem was defined by SAP and Ariba [30]. The Open Catalog Interface (OCI) or 'punch-out' solution offers a purchaser the possibility to see the catalogue on a webpage formatted in html. The catalogue does not enter the ERP system of the purchaser. The purchaser can select the products he wants to order in the web interface, and he can have a standardized message sent to his ERP system so that (only) the data on the desired products enters his ERP system.

We just stated that the suppliers may not want the purchasers to load the entire catalogue into their system because this would make it too easy for the customers to compare prices with competitors. However, if the competitive strength of the company is just that it *has* the lowest prices, it *will* try to make its pricing data as accessible as possible. In case the supplier tries to differentiate his products from those of competitors using other variables than the price, the supplier will try to pull the attention of the customers to the appropriate variables; for example by presenting a nice picture on the catalogue-webpage. That is, in case of Market B2Bi it would work that way. In the case of Extended Enterprise integration, prices *could* be made available in an accessible format, even if they do not show the best side of the company. Still, this offers the partner the possibility to poach (see later).

ILLUSTRATIONS

If one company uses an SAP ERP system and another company uses an Oracle ERP system, a solution must be searched to enable both systems to talk with each other. Nowadays Commercial-Off-the-Shelf software packages often support a number of standards so that a common format to exchange data between the packages of different vendors is readily available (enabling a best-of-breed approach).

In the context of VMI a number of standard EDI messages are typically used (such as messages 852, 855 and 856). These EDI messages are a standard data transmission format. The data that is stored in the systems of the parties then has to be translated to this EDI format for transmission. Although these EDI messages are standardized it is still important to test the EDI messages extensively before taking VMI to the operational stage [2]. The seller and the reseller have to work closely together to validate the messages: does some message result in a correct representation of the state at the seller's site? For example, is the Quantity On Hand being communicated correctly? That is, is the message built correctly at the reseller's site, and is the message 'consumed' correctly at the seller's site?

When considering PLM, we see that for designers a dumb data file (i.e., raster data which are essentially bit map pictures) is not very useful: they need the vector formats to be able to make changes to the product's design. The other way around, users do not want to make changes to a product's design, and a vector representation of the product in the users' user-manual is not desirable.

Different parties have to make investments

To realize the information sharing, investments will need to be made. Partners have to agree on who will bear the costs for installing, maintaining and upgrading the systems.

PROBLEM DESCRIPTION

Investments can be left primarily to one partner or can be split among the partners and/or third parties. If one party makes changes to (or withdraws) the systems he possesses, other parties may need to make investments.

RELATIONSHIP TO THE CHALLENGE

To make sure the data is available to the parties it is desirable that new parties that want to enter the network do not need to make much investments, and that parties that want to leave do not take much of the infrastructure with them so that the information sharing for the remaining parties remains operational. Also, it is possible that some information flows can no longer be executed if one party changes its investments.

RELEVANCE

If a party makes changes to his own investments, for example by upgrading his systems, he may cause some information flows to fail (e.g., because some interfaces become invalid). As partners in an Extended Enterprise know who may use the service they can communicate (and negotiate) the changes to those parties.

If parties can change over time the party that leaves can take some part of the investments with him so that the rest of the Extended Enterprise suffers more than proportionally from the departure of that party (i.e., they lose more than just the connection with that party: they also lose connection with other parties).

A new partner may have to make many investments himself to get into the network. This makes the decision for him to enter the network harder as it possibly results in a lock-in that may not be desirable at the beginning of a partnership. Also, companies of which it is interesting to get information may not be willing to give that information

because they fear that the information will be exploited (i.e., be used beyond a contractual relationship) if the data sharing system is no neutral property [31]. This seems especially true in the case of coopetition (i.e., cooperation between competitors). It is thus not only important to decide who pays for the investment just because it *directly* influences who wants to share data (i.e., as it costs money you do *not* want to share data), but also because it *indirectly* has an influence (i.e., as you own the storage space yourself you *are* willing to share data).

ILLUSTRATIONS

The investment-distribution depends upon the chosen information sharing solution. Therefore, we cannot give a general comment on VMI or PLM. We can, however, shortly look at the Tradcom-case. There, customers do not have to make big investments: they only have to enable their systems to send purchase orders in their proprietary format to Tradcom. The suppliers have to make bigger investments: they need to communicate in a Tradcom-specific XML format with Tradcom. Tradcom translates the messages from the customers to this Tradcom-specific format. A group of suppliers together own Tradcom and they are thus the parties that invest in the communication. Customers can easily enter and leave the marketplace while suppliers are closely tied to Tradcom. Tradcom lowers the investment for the suppliers because it is easier for suppliers to connect only to Tradcom than having to connect to all systems of all customers. If a new customer enters the marketplace, the individual suppliers do not have to make any investments. If a new supplier enters the marketplace, other suppliers do not have to make any investments and neither do the customers. The question is what would happen if one or several of the Tradcom-owning suppliers would ‘leave’ Tradcom. If they would take the Tradcom platform with them, the customers and all other suppliers would be disconnected.

Partners become dependent upon the service levels provided by the data sharing systems

If data is to be valuable for some company it has to be offered with appropriate service levels.

PROBLEM DESCRIPTION

The most important service level requirement concerns ‘availability’. Partners need data to execute desired tasks, but there is a chance that at some point in time the data will be unavailable, hampering the internal functioning of one or more companies and/or the functioning of the entire Extended Enterprise.

The fact that data may not be available of course creates problems for the business. Still, the availability problem reaches out further. Partners cannot (expect each other to) provide a 99.99% availability. The higher the required availability, the higher the cost for meeting that requirement. An additional problem is thus that the service level expectations of the users should be managed so that providing a lower availability does not create negative feelings.

Other service level issues that may require attention are the time span that backups are saved, the response time of the systems, the variability in response time, the maximum time span the data may not be accessible, etc. We note that current solutions to deal with contingencies (e.g., by calling your colleague in the partnering company) may become less evident after some time. This is because Business-to-Business systems integration may weaken the personal relationships that exist between companies [26], increasing the dependence upon the well-functioning of the computer systems.

RELATIONSHIP TO THE CHALLENGE

The data the partners are *expected* to have access to may not be available at some point in time, or its availability may be corrupted by poor service levels.

RELEVANCE

Tasks cannot be executed as desired and the supply chain cannot function optimally if at some moment the necessary information would not be available (especially if the problem persists). Partners may therefore set up Service Level Agreements (SLAs).

In terms of Levina and Vaast [26] partnering companies function like a 'community'. In [26, 29] it is shown that the use of ICT may help deteriorate community ties, turning the community into a 'market'[2]. One reason for this is that the subgroups of the community that were tied together by the boundary spanners fall apart because the boundary spanners are made obsolete. If the 'objectification' does not decently replace the 'embodiment' of the personal relations, this may be disastrous (as is the case in [26]). Once the personal relationships are dropped it may be hard to make them operational again.

ILLUSTRATIONS

In the context of VMI: if the reseller's stock and sales data for some reason are not available to the seller, the latter is not able to accurately define the 'purchase orders' (i.e., 'sales orders' from his point of view) he should create. The network may be able to function without the information, but *only for some time*.

Similarly, in the case of shared product data companies most likely can execute their tasks if they have access to not-so-recent product data. Still, sometimes it will be important that an update *is* considered in the execution of a task. The infrastructure should then allow the timely communication of this update.

Product data files can be quite large (e.g. > 100 MB). If such a file is 'suddenly' needed the requesting system may have to wait quite some time for the transmission to complete, disrupting the system's execution.

Partners must preserve the value of the functional proposition of the data sharing

The partners should handle the data in line with how other partners would like them to handle it so that they can be trusted.

PROBLEM DESCRIPTION

We consider two main issues on which companies have to trust each other: the receiver has faith in the quality of the data, and the sender has faith in the receiver preserving the confidentiality of the data. This trust can be harmed directly (by the partner himself), or indirectly (by a third party). This is shown in Figure 1.

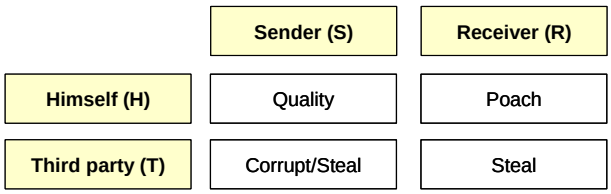

Figure 1: Sender and receiver are both responsible for preserving trust

[2] Market practices are here defined as practices that involve 'an exchange and combination of work outcomes that are, for the most part, produced separately' by the different agents, rather than jointly derived [26, p18].

We note that in other literature it is said that in other cases ICT may help *build* community ties (see [32] for example).

Let us first consider the receiver's responsibilities. Partners that get new data at their disposal have the power to 'poach'. Poaching can be defined as involving three components: '(1) the exchange of information between two parties, as a natural byproduct of contractual exchange for other goods or services, necessary for the performance of contractual obligation; (2) the subsequent use of this information by the receiving party, outside the purposes for which the information was provided, and for its own benefit or economic gain; and (3) at the expense of, or creating economic damage to, the party that provided the information' [28, p94]. If we extend this definition, poaching cannot only concern 'production data', but also meta-data: partners may pass on data about what data is being shared.

Going one step further, the receiving party may reject poaching but inadequately secure its systems so that unauthorized parties can access the data. The idea arises that a network is only as secure as its weakest link. Security requires authentication and authorization. We note that while authorizations are typically discussed in the context of *who* is trying to access some data content it is important to look further. More specifically, authorizations may depend on what task the data will be used for, the physical location the message has to be sent to, the medium over which the message is transmitted, the moment the message is sent, the data content, the message format, whether it concerns a single record or a batch of records, whether the data is coarse-grained or fine-grained, etc. [33]. Authorizations are thus a complex matter.

Of course, it is not only the receiving system that should be trustworthy. The sending party should make sure that the data it provides is of the expected quality. Therefore, this party should not only prevent corruption (by internal or external individuals) of stored data but should also make sure the data is entered in an appropriate quality. This may be a heavy task, as the operators may have to enter data of which they do not know or understand for what purpose it could be relevant (for the other party).

A special problem concerning the preservation of the functionality of the data has to do with non-repudiation. If a transmission is successful it will most likely have consequences (e.g., some party will execute a task on the request of some other party). These consequences are very real, and a party that has taken part in the transmission of information should not disaffirm its participation later on. Otherwise later transmissions may lose their value. In a B2B context, the concept of non-repudiation is important. This concept embraces two ideas: the sender cannot deny that he sent the message [34], and the receiver cannot deny that he received the message [20].

Relationship to the challenge

The data a party gets should be of an appropriate quality, and only authorized parties should get access to the data.

Relevance

'Privacy' is a topic that is often mentioned in the context of partnering companies. In general, information privacy concerns the fact that individuals require that information about themselves should generally not be available to others, and that, where data is possessed by another party, the individual should be able to control the data and its use to a considerable extent [35]. From this we can say that if a customer gives personal information to a specific company, this does not mean the customer (the owner of the data, see below) would agree to give this information to the partners of this company. The partners should clearly handle the data with care, fully respecting the agreement the originating organization has with its customer.

This partner may not have a direct relationship with that customer, elevating chances of

poaching. One would not expect poaching to show up in an Extended Enterprise context as poaching would most likely damage trust. Still, actual poaching is difficult to observe [28], turning it more plausible. Also, poaching is more likely to turn up if there is a weak intellectual property protection (e.g., the impossibility to 'return' the information at the end of the contract) and if poaching was not prohibited in a contract [28], two conditions that may apply in a close partnership.
One could make a similar assumption about the problem of securing the systems: one could assume that partners decently secure their systems as they know they are dealing with confidential information. In one study, Dynes [37] indeed found that firms do not formulate big security requirements for their suppliers. Yet, in another study he found that in general companies *are* auditing the information security status of potential partners. One difficulty with such assessments is that they are said to slow down the partnering [38].
Guaranteeing the quality of the data can happen more decently in an Extended Enterprise than in Market B2Bi. The data quality is not just a given; data is not just made available 'as-is' to be used by a multitude of counterparties. Partners can negotiate the importance of the 'validity' of the data (and what is 'valid' to them), so as to invest an appropriate amount of energy in making and keeping the data valid.
Taking care of non-repudiation is considered to be more relevant in a context of Market B2Bi than in Extended Enterprise integration.

ILLUSTRATIONS
If we apply Figure 1 to the PLM example:
(RH) The party may consciously share his partner's designs with a competitor of that partner.
(RT) If some party's systems are not well secured, intruders may steal the product designs that are owned by this party's partner.
(SH) For one party it may suffice that data are exact to the centimetre, while for another party millimetres may be relevant. The rounding that the former party (implicitly) performs results in inaccurate data for the latter party.
(ST) If some external party can hack the systems and can change some measurements, the plans become worthless. If these kinds of problems re-occur this makes it hard for companies to have confidence in the data that is being shared.
Similar examples could be given for the VMI case. For example, with VMI it is important for the seller to get accurate sales data, and thus no data that is inflated to reassure stockholders. For resellers it is important that the seller secures his systems so that competitors cannot find out about a promotional campaign that he planned.

Data ownership may not be well arranged

PROBLEM DESCRIPTION
On the one hand, *several* parties may want to decide what can/must happen with the data. On the other hand, *no* party may be designated as data owner or may feel responsible.
Data ownership has been defined as the "responsibility for determining the required quality of the data, for establishing security and privacy for the data and determining the availability and performance requirements for the data". It is all about "data originators who have the authority, accountability, and responsibility to create and enforce organizational rules and policies for business data" [39].

RELATIONSHIP TO THE CHALLENGE

Who has the responsibility to determine who can access the data (and who cannot) with what service levels and who has to make the data valid?

RELEVANCE

If ownership is not well arranged, data that is considered to be valid may in fact not be valid; data may fall in the wrong hands, etc.

In [40] it is stated that one of the most critical legal risks that confront Extended Enterprises is the ownership of intellectual property. In Market B2Bi each party is likely to own its own view on the object. In market-like practices 'objects are separately produced and exchanged according to prespecified terms' [26, p28]. In an Extended Enterprise, different partners may collaboratively change some view on the object so that it may become obscure who actually owns some view on the object. In community-like practices 'boundary objects are coproduced, they represent negotiated outcomes of a joint practice' [26, p28].

ILLUSTRATIONS

If a subpart of a product is created by a supplier, who is then the owner of the subpart-data? Is it the producer of the final product, or the producer (presumably the designer) of the subpart? In case of Market B2Bi this is most likely the producer of the subpart, but in case of Extended Enterprise integration this may be less clear. Depending on who is the owner, changing partners may result in changing the ownership of parts of the BOM data.

The problem may even be more clearly illustrated with VMI. Who is the owner of the stock data in this case? Unless consignment is used, the stock is the property of the reseller. This would show the reseller as the data owner. The stock is, however, managed by the seller. So, the seller may be seen as the data owner. The fact that the reseller and the seller have to agree on a stocking plan, inventory turns, fill rates, replenishment frequency and the like shows that both parties have something to say about the stock. In specific cases one party may be more powerful than another one, determining the data ownership.

While the previous examples concern data that 'exist at the level of individual enterprises', data ownership problems also show up in cases where there are data that only exist at the level of the entire Extended Enterprise. We can illustrate this with the Tradcom-case. The suppliers sell their products through Tradcom to a multitude of customers. While individual suppliers only get orders with respect to their own products, there is information avalaible at Tradcom-level about orders that contain products from different suppliers, and about customers that entered the platform to do business with one supplier but now also do business with other suppliers, etcetera. That is, new information is available at the Extended Enterprise level that was not previously available at the level of the individual companies. The question is who owns this data and is responsible for managing its quality. This information can be very valuable but can easily be overlooked because it is by nature not really owned by any of the individual parties.

The involved parties change over time

PROBLEM DESCRIPTION

The parties that *provide* the data, and/or the parties that *use* the data may change. There is a risk that partners use old, invalid data and that former partners still get access to data present in the Extended Enterprise.

RELATIONSHIP TO THE CHALLENGE

The data that is *valid* changes if partners change. Also, there are changes in the *agreement* on which parties can access data: some parties can probably no longer access the data. This decision has to be assessed and monitored.

RELEVANCE

Partners in an Extended Enterprise are not likely to be dropped from the network. That is not to say that partners never may leave or be added to the network. Partners that are dropped from the Extended Enterprise are likely to become competitors of the Extended Enterprise. If they still have access to the Extended Enterprise data this is likely to be baleful for the competitive position of the Extended Enterprise. However, sometimes it is normal (and desirable) that they *do* get access to some of the data, for example to offer after-sales service. Similarly, it is often important for members of the Extended Enterprise to still have access to the data of the former partner.

An entrant should be able to get access to data from partners and should be able to share his data with partners in order to become really part of the Extended Enterprise.

ILLUSTRATIONS

If the supplier of a semifinished product is replaced, the product data set that is valid changes. That is, in the design of the assembled product, the design of the semifinished product has to be replaced by the design of that part made by the new provider (or at least its interface). While the old design data may not be valid for newly produced products it remains useful for maintaining old products and offering after-sales service. This implies that different versions of the component design have to be managed and related to specific instances of the produced product.

In general, the former supplier of the component should no longer have access to the other product data as he might spread this data across the new network he entered more recently. Nevertheless, if this supplier is still to create the spare-parts for the product under consideration he may need to be given access to some part of the data.

The fact that the supplier can typically take the data on the component he manufactures with him is usually not considered a problem, even if the partners in the Extended Enterprise helped him to create an improved design. After all, that design is very likely to be completely aligned with the design of the other components of the product, and competing Extended Enterprises are less likely to benefit from that improved design as there is a big chance on misalignment with their designs and their way of working. In the Toyota Extended Enterprise for example, “Toyota is able to generate competitive advantages through knowledge sharing in its extended enterprise even when it uses the identical suppliers as its competitors” [41, p85].

With VMI the seller truly obtains insight in the reseller’s business. This can be illustrated with the fact that if a transmission of sales and stock data would fail, the seller – in practice – would still have quite a good idea of the sales and stock levels at the reseller’s site because he knows his business so well. This fact is very desirable for as long as both companies do business with each other, but once their relationship falls apart the seller may try to use this knowledge in his relationship with new resellers.

CONCLUSIONS

This paper discussed eight problems that show up when partnering companies decide to share information. Each of the problems was shown to occur in different practical situations. We used two different situations. One concerned the PLM industry where

the shared files are likely to be large and transmitted sporadically. The other one concerned VMI. Transactional data records (about sales and stocks) are less voluminous. For both cases the problems could be instantiated.

It is important in practice to deal with each of these problems. Our research (not reported in this paper) showed that different Business-to-Business integration solutions have different capabilities to deal with each of the problems. For example, if one considers the 'boundary object' problem, one has to recognize that it is typical for centrally controlled systems to have difficulties in capturing local understandings, complicating the creation of an Extended Enterprise wide data model. From this, a centralized B2Bi solution seems most appropriate in case one is dealing with standard data structures and data mappings. For the PLM industry, standard data schemas have been developed in the past, and could be used to build a centralized repository, as was done by Eurostep in it's Share-A-Space solution for example. In case the mapping between data of different parties is not standardized it becomes very difficult to build such an n-to-n mapping at the centralized level.

Besides the 'boundary object' problem, the problem that investments and benefits need to be distributed was discussed. The amount of counterparty-specific investments companies want to make depends upon the (expected) duration of the use of the investment and thus upon the (expected) duration of the relationship between the companies. In an Extended Enterprise this duration is typically (very) long. That is not to say that new partners may not be added to the Extended Enterprise, and that current partners never may leave it.

We noted that a data-owner is needed at inter-organizational level which determines who can access what data and who cannot. Furthermore, it was stated that the value of the information sharing solution depends on the service levels that are provided. Partners are, however, not only dependent upon each other for offering appropriate service levels, they are also contingent upon the way the other handles the data. The sender is dependent upon his partners to reject poaching, and to secure their systems. The receivers are contingent upon the senders for transmitting the data in an appropriate quality. Above that, problems with respect to the data format and unclarity about the appropriate data flows may complicate things.

REFERENCES

[1] Smaros J. Lehtonen J.M., Appelqvist P., & Holmström J. (2003). The impact of increasing demand visibility on production and inventory control efficiency. *International journal of physical distribution & logistics management*. Vol. 33, No. 4, pp. 336-354.

[2] VendorManagedInventroy.com, Vendor Managed Inventory Set Up. Retrieved from http://www.vendormanagedinventory.com/setup.htm on April 4, 2006.

[3] deMin J.E. (2003). Collaborative Planning, Forecasting and Replenishment (CPFR) and the Network – Cracking the Bullwhip! Retrieved from http://wp.bitpipe.com/ on September 28, 2006, p 17.

[4] Takalikar S. (2004). Leveraging PLM for manufacturing excellence, *Express computer*, April 2004, Retrieved from http://www.expresscomputeronline.com/ on April 5, 2006.

[5] CIMdata Inc. What is PLM?, Retrieved from http://www.cimdata.com/PLM/aboutPLM.html on November 4, 2005.

[6] The PLM Alliance, Reconciliation of multiple bills of materials, Retrieved from http://plm.engin.umich.edu/BOMReconciliation.htm on April 11, 2006.

[7] Moeller G. (2000). Standard for the Exchange of Product Model Data (STEP); PM, November-december 2000, pp. 52-56.
[8] Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science* 5:1, pp. 14-37.
[9] Inmon B. (2004). Looking Ahead: Unstructured Data, *DM Review*, August 2004, p. 56.
[10] Drucker P.F., Garvin D., Leonard D., Straus S., & Brown J.S. (1998). *Harvard Business Review on Knowledge Management*. Harvard Business School Press. p 223.
[11] Isaacs E., Walendowski A., Whittaker S., Schiano D., & Kamm C. (2002). "The Character, Functions, and Styles of Instant Messaging in the Workplace. *Proceedings of the CSCW'02*. 126-135. New Orleans, Louisiana, 2002.
[12] Ribak A., Jacovi M., & Soraka V. (2002). Ask Before You Search: Peer Support and Community Building with ReachOut. *Proceedings of the CSCW'02*, 126–135. New Orleans, La.
[13] Woo J.H., Clayton M., Johnson R., Flores B., & Ellis C. (2005). Sharing tacit design knowledge in a distributed design environment, *AIA Report on University Research*, pp. 47-58, Retrieved from http://www.aia.org/ on September 15, 2005.
[14] Kim T-G., Yu S-H., Lee J-W. (2003). Knowledge strategy planning: methodology and case. Expert systems with applications, 24: 295-307.
[15] Polanyi M. (1966). *The tacit dimension*, Routledge & Kegan Paul ltd. London.
[16] Johansson M. White paper: Sharing product data through life in the extended enterprise – the Share-A-Space solution, p. 7.
[17] Sheu R.K., Chang Y.S., & Yuan S.M. (2001), Managing and sharing collaborative files through WWW, Future Generation Computer Systems, 17:8, pp. 1039-1049.
[18] Tang M., Lee B.S., Tang X, & Yeo C.K (2006). The impact of data replication on job scheduling performance in the Data Grid, Future Generation Computer Systems, 22:3, pp. 254-268.
[19] Fatteross J. Release-to-manufacturing Without the Back-and-Forth, Retrieved from http://www.imakenews.com/ On April 6, 2006.
[20] http://www.unesco.org/webworld/mdm/administ/en/guide/guide011.htm
[21] Kumar P., & Kumar M. (2003). Vendor Managed Inventory in Retail Industry. Retrieved from http://www.tcs.com/ on May 24, 2006.
[22] Carlile P.R. (2002). A pragmatic view of knowledge and boundaries: boundary objects in New Product Development, *Organization Science*, Vol. 13, No. 4, pp. 442-455.
[23] Star S.L., & Griesemer J.R. (1989). Institutional Ecology, Translations and Boundary objects: amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-39, *Social Studies of Science* (SAGE, London, Newbury Park and New Delhi), Vol. 19, pp. 387-420.
[24] Levina N., & Vaast E. (2005). The emergence of boundary spanning competence in practice: implications for implementation and use of information systems. *MIS Quarterly*, Vol. 29, No. 2, June 2005, pp. 335-363.
[25] Snoeck M, Lemahieu W, Goethals F, Dedene G, Vandenbulcke J. (2004). Events as atomic contracts for application integration. *Data and Knowledge Engineering* 2004 Oct; 51(1), pp.81-107.

[26] Levina N., & Vaast E. (2006). Turning a community into a market: a practice perspective on information technology use in boundary spanning. Journal of Management Information Systems. Vol. 22, No. 4, pp. 13-37.
[27] Caronna S., Data granularity in the storage and reporting of soil exploration information, Retrieved from http://www.gintsoftware.com.
[28] Clemons E.K., & Hitt L.M., (2004). Poaching and the misappropriation of information: transaction risks of information exchange. *Journal of Management Information Systems*, Vol. 21, No. 2, pp. 87-107.
[29] Schultze, U., Orlikowski W.J. (2004). A practice perspective on technology-mediated network relations: the use of internet-based self-serve technologies. Information Systems Research, Vol. 15, No. 1, pp. 87-106.
[30] SAP (2003), Open Catalog Interface (OCI) Release 4.0. Retrieved from https://www.sdn.sap.com/irj/servlet/prt/portal/prtroot/com.sap.km.cm.docs/library/icc/Open%20Catalog%20Interface.pdf on July 20, 2005.
[31] Han K., Kauffman R.J., & Nault B.R. (2004). Information exploitation and interorganizational system ownership, *Journal of Management Information Systems*, fall 2004, Vol. 21. No. 2, pp. 109-135.
[32] Levina N., & Vaast E. (2005). The emergence of boundary spanning competence in practice: implications for implementation and use of information systems. *MIS Quarterly*, Vol. 29, No. 2, pp. 335-363.
[33] Goethals F., Lemahieu W., Snoeck M., & Vandenbulcke J. (2006). Dependencies between Data Decisions, Software Process Change - *Proceedings of the SPW/ProSim Workshop*, Shanghai, May 20-21, 2006, LNCS Volume 3966, pp. 48-55.
[34] DigitalGuards, Glossary, Retrieved from http://www.tsl.state.tx.us/ on November 12, 2006.
[35] Clarke, R. (1999). Internet Privacy Concern Confirm the Case for Intervention. *Communications of the ACM*, 42, 2, (February 1999), pp. 60-67.
[36] Fürst K., Schmidt T., Wippel G. (2002), Managing Access in Extended Enterprise Networks, *IEEE Internet Computing*, September-Oktober 2002, pp. 67-74.
[37] Dynes S., Brechbüchl H. Johnson M.E. (2005). Information security in the extended enterprise.
[38] Dynes S. Security and Privacy: at odds with speed and collaboration?, p. 13.
[39] California State University, Data Warehouse Glossary, Retrieved from http://it.csumb.edu/departments/data/glossary.html on June 19, 2006.
[40] Knowledge @ Wharton. Special section: The rush to send back-office business overseas – Part 2, Retrieved from http://knowledge.wharton.upenn.edu on April 25, 2006.
[41] Dyer J.H. (2000). Collaborative Advantage – Winning through extended enterprise supplier networks, Oxford University Press, p. 209.